NP04 10G performance issues

Roland Sipos

CoreSW Meeting
26th June 2024

Problem statement

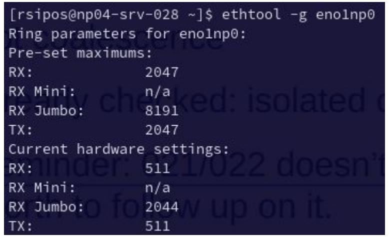
- High trigger rate issues (30+ Hz)
  - Causes back pressure on the readout data reception (missed packets), but interestingly only on the two less powerful servers: 021/022
  - Trigger gets inhibited
  - With TPG on, the situation is even worse
  - Other errors or affected subsystems and components?

Software

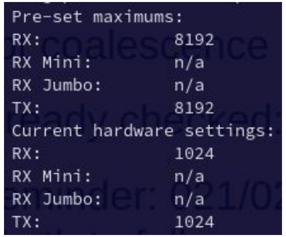
- We can start with some configurable and obvious checks and tests
- Readout
  - Number of request response threads can be increased. (Default: 4, Go for 10?)
- Dataflow
  - Are there test applications for 10G link saturation? I think we can easily configure an emulated system to see how far can we push the link.
- Appfwk
  - We have IOManager tests, that could be used for link saturation tests, and we should definitely make a scan with different payload sizes
  - ZMQ doesn’t have many tweaks, only suggestions for increasing file descriptor limits

Hardware - 10G NICs

- This is system administration work, but we must aid with the testing and evaluation
- Are there dropped packets? (Didn’t see on the readout, to be checked on storage)
  - Modify ring buffers count to hardware limits
    - Defaults everywhere at the moment.



```
[rsipos@np04-srv-028 ~]$ ethtool -g eno1np0
Ring parameters for eno1np0:
Pre-set maximums:
RX:             2047
RX Mini:        n/a
RX Jumbo:       8191
TX:             2047
Current hardware settings:
RX:             511
RX Mini:        n/a
RX Jumbo:       2044
TX:             511
```



```
Pre-set maximums:
RX:             8192
RX Mini:        n/a
RX Jumbo:       n/a
TX:             8192
Current hardware settings:
RX:             1024
RX Mini:        n/a
RX Jumbo:       n/a
TX:             1024
```

  - Interrupt coalescence
    - Already checked: isolated cores are free from IRQs! This is very good news.
    - Reminder: 021/022 doesn’t isolate TPG threads’ CPUs! There were some issues. Worth to follow up on it.
- Jumbo frames (MTU 9000)
  - Either all servers or none -> effect on the whole subnet

Kernel

- This is system administration work, but we must aid with the testing and evaluation
- Many possible options… quite overwhelming to test everything. We should focus on the first and immediate things, like:
    - Tuned-adm profile is currently on latency-performance (?? should try network ones)
    - Tuning for throughput or latency? There are certain options that will benefit only either of one
- We should really apply some obvious parameters from the Linux TCP tuning guide that is discussed here:

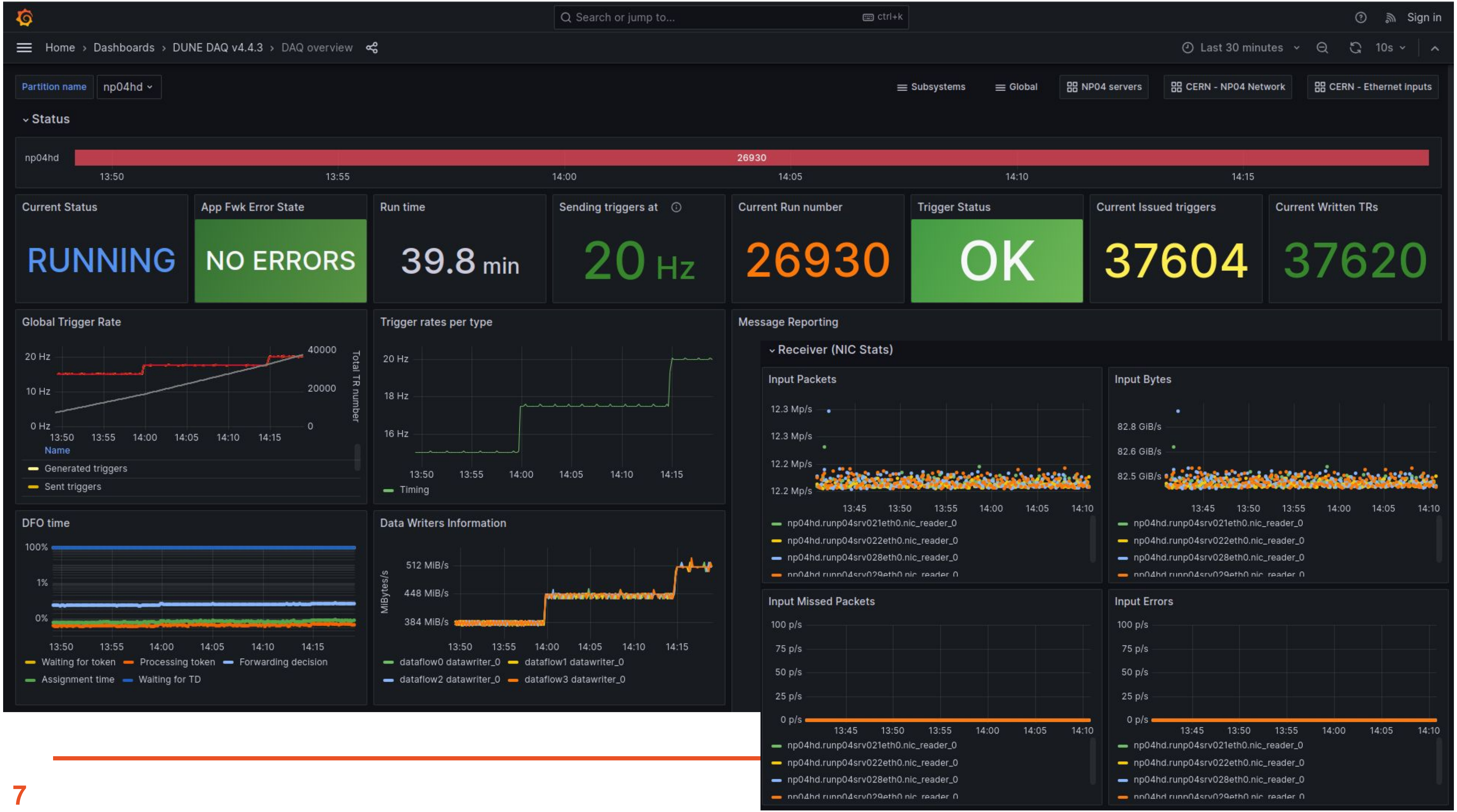
  http://www.linux-admins.net/2010/09/linux-tcp-tuning.html

  Best examples: TCP buffer sizes, netdev max backlog

Approach

There are many options to look into and evaluate if the settings are beneficial. Tweaking too many parameters in one go will lead to a mixed understanding of the results.

- Change one thing at a time, we need to be systematic!



Home › Dashboards › DUNE DAQ v4.4.3 › DAQ overview

Last 30 minutes | 10s

Partition name: np04hd

Subsystems | Global | NP04 servers | CERN - NP04 Network | CERN - Ethernet inputs

Status

np04hd: 26930

Current Status: RUNNING

App Fwk Error State: NO ERRORS

Run time: 39.8 min

Sending triggers at: 20 Hz

Current Run number: 26930

Trigger Status: OK

Current Issued triggers: 37604

Current Written TRs: 37620

Message Reporting

Receiver (NIC Stats)